09/509196

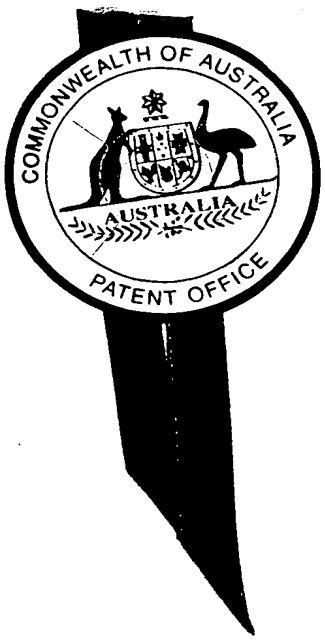
PCT/AU98/00795

Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PO 9388 for a patent by GARVAN INSTITUTE OF MEDICAL RESEARCH filed on 23 September 1997.

I further certify that the annexed specification is not, as yet, open to public inspection.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)



WITNESS my hand this Sixth day of October 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

GARVAN INSTITUTE OF MEDICAL RESEARCH

PROVISIONAL SPECIFICATION

Invention Title:

A potential effector for the Grb7 family of signalling proteins.

The invention is described in the following statement:

The present invention relates to a novel polynucleotide molecule encoding a polypeptide which represents a candidate effector protein for the Grb7 family of signalling proteins. Detection of the encoded protein in a tissue sample may provide a useful tumour marker and/or prognostic indicator. Furthermore, antagonism of the interaction between Grb7 family members and this effector protein may provide a novel treatment strategy for human diseases exhibiting aberrant receptor tyrosine kinase (RTK) signalling (e.g. cancer).

RTKs play a major role in the regulation of cellular growth, differentiation, motility and metabolism by converting an extracellular signal in the form of the binding of a specific hormone or growth factor to the activation of specific signalling pathways and hence modes of intracellular communication (Schlessinger and Ullrich, Neuron 9, 383-391, 1992). Activation of RTKs results in both autophosphorylation of the receptor and the phosphorylation of downstream targets on tyrosine residues. It has become evident over the last decade that key elements in receptor-substrate and other protein-protein interactions in RTK signalling are src homology (SH)2 domains. SH2 domains are conserved modules of approximately 100 amino acids found in a wide variety of signalling molecules which bind to short tyrosine-phosphorylated peptide sequences. The specificity of interaction is determined both by the nature of the amino acids flanking the phosphotyrosine residue in the target peptide and residues in the SH2 domain which interact with these sites (Pawson, Nature 373, 573-580, 1995).

SH2-domain containing proteins can be divided into two classes: those which possess a catalytic function (eg the cytoplasmic tyrosine kinase c-src and the tyrosine phosphatase SH-PTP2) and those which consist entirely of non-catalytic protein domains (eg Grb2), the adaptor sub-class. The function of the latter class is to link separate catalytic subunits to a tyrosine-phosphorylated receptor or signalling intermediate, and other non-catalytic protein modules are often involved in these interactions. For example, SH3 and WW domains (conserved regions of approximately 50 and 40 amino acids, respectively) bind proline-rich peptide ligands, and pleckstrin homology domains (approximately 100 amino acids) interact with both specific phospholipid and protein targets (Pawson, 1995).

The Grb7 family represents a family of SH2 domain-containing adaptors which currently contains three members: Grb7, 10 and 14 (Margolis *et al*, Proc. Natl. Acad. Sci. USA 89, 8894-8898, 1992; Stein *et al*, EMBO J 13, 1331-1340, 1994; Ooi *et al*, Oncogene 10, 1621-1630, 1995; Daly *et al*, J. Biol. Chem. 271, 12502-12510, 1996). These proteins share a common overall architecture, consisting of an N-terminal region containing a highly conserved proline-rich decapeptide motif, a central region harbouring a PH domain and a C-terminal SH2 domain. The central region of approximately 300 amino acids bears significant homology to the *C. elegans* protein mig10, which is required for long range neuronal migration in embryos, otherwise the Grb7 family and mig10 are structurally distinct. However, they exhibit differences in both SH2 selectivity towards RTKs (Janes *et al*, J. Biol. Chem. 272, 8490-8497, 1997) and tissue distribution. The family has therefore evolved to link particular receptors to downstream effectors in a tissue-specific manner. Interestingly, the genes encoding this family appear to have co-segregated with *ERBB* family genes during evolution. Thus *GRB7, 10* and *14* are linked to *ERBB2, ERBB1* (epidermal growth factor receptor) and *ERBB4*, respectively (Stein *et al* 1994; Ooi *et al*, 1995, Baker *et al*, Genomics 36, 218-220, 1996). The juxtaposition of *GRB7* and *ERBB2* leads to common co-amplification in human breast cancers, and since the two gene products are functionally linked, likely up-regulation of an undefined erbB2 signalling pathway. Furthermore, *GRB14* also exhibits differential expression in human breast cancers (Daly *et al*, 1996). These two proteins may therefore modulate RTK signalling in this disease.

In order to identify proteins which bind to this family and therefore identify candidate effectors, we performed a genetic screen using the yeast two hybrid system and Grb14 "bait". This application describes the cloning and characterization of a novel interacting protein, currently designated 2.2412.

Thus, in a first aspect, the present invention provides an isolated polynucleotide molecule encoding a candidate effector protein for the Grb7 family of signalling proteins, wherein the polynucleotide molecule comprises a nucleotide sequence encoding a polypeptide comprising an amino acid sequence substantially as shown in Figure 1.

In a preferred embodiment of the present invention, the polynucleotide molecule comprises a nucleotide sequence substantially as shown in Figure 1.

In a second aspect, the present invention provides a purified polypeptide being encoded by a polynucleotide molecule which hybridises to the nucleotide sequence of Figure 1 under medium to high stringency conditions (Sambrook *et al*, Molecular Cloning: a laboratory manual, Second Edition, Cold Spring Harbor Laboratory Press).

The polynucleotide molecule referred to in the first and second aspects, may be a dominant negative mutant which encodes a gene product causing an altered phenotype by, for example, reducing or eliminatingthe activity of endogenous effector proteins of the Grb7 family of signalling proteins.

In a preferred embodiment of this aspect, the purified polypeptide comprises an amino acid sequence substantially as shown in Figure 1.

In a third aspect, the present invention provides a fusion protein comprising an amino acid sequence substantially as shown in Figure 1.

Fusion proteins according to the third aspect may include a N-terminal fragment of a protein such as ϑGal to assist in the expression and selection of host cells expressing candidate effector protein, or may include a functional fragment of any other suitable protein to confer additional activity(ies).

In a fourth aspect, the present invention provides an antibody which binds to the polypeptide of the second aspect.

The antibody may be monoclonal or polyclonal, however, it is presently preferred that the antibody is a monoclonal antibody.

In a fifth aspect, the present invention provides an oligonucleotide probe of at least 12 nucleotides, the oligonucleotide comprising a nucleotide sequence such that the oligonucleotide probe selectively hybridises to the polynucleotide molecule of the first aspect under high stringency conditions.

In a preferred embodiment of this aspect, the oligonucleotide probe is labelled. In a further preferred embodiment of this aspect, the oligonucleotide probe is of at least 18 nucleotides.

In a sixth aspect, the present invention provides a method of detecting in a sample the presence of an effector protein for the Grb7 family of proteins, the method comprising reacting the sample with an antibody of

the fourth aspect or an oligonucleotide probe of the fifth aspect, and detecting the binding of the antibody or the probe.

The terms "comprise", "comprises" and "comprising" as used throughout the specification are intended to refer to the inclusion of a stated component or feature of group of components of features with or without the inclusion of a further component or feature or group of components or features.

In order that the nature of the present invention may be more clearly understood, preferred forms thereof will now be described with reference to the following example and Figures in which;-

Figure 1 provides the nucleotide and amino acid sequence of cDNA clone 2.2412. The amino acid sequence (single letter code) is indicated in boldface type. Numbers refer to distances in base pairs.

Figure 2 provides an analysis of the amino acid sequence of cDNA clone 2.2412. Ankyrin-type repeat sequences are indicated in boldface type. Each repeat is numbered consecutively above the first residue. An additional repeat sequence is underlined. Amino acid residues are numbered on the left of the figure.

EXAMPLE

CLONING AND CHARACTERISATION OF 2.2412

Yeast two hybrid screen

The yeast two hybrid system exploits protein-protein interactions to reconstitute a functional transcriptional activator which can then be detected using a gene reporter system (Fields and Sternglanz, TIG, 10, 286-292, 1994). The technique takes advantage of the properties of the Gal4 protein of the yeast *S. cerevisiae*. The Gal4 DNA binding domain (DNA-BD) or activation domain (AD) alone are incapable of inducing transcription. However, an interaction between two proteins synthesized as DNA-BD- and AD-fusions, respectively, brings the Gal4 domains into close proximity and results in transcriptional activation of two reporter genes (*HIS3* and *LacZ*) which can be monitored by growth on selective medium and biochemical assays.

A plasmid construct encoding a Gal4 DNA-BD-Grb14 fusion was generated as follows. The plasmid *GRB14*/pRcCMVF containing full length *GRB14* cDNA (Daly *et al*, 1996) was restricted with HindIII and Klenow treated to create blunt ends, and then digested with BclI to release three fragments of approximately 1.1, 4.2 and 1.7 kb. The 1.7 kb fragment was isolated and cloned into the NdeI (Klenow treated) and BamHI sites of the yeast expression vector pAS2.1 (Clontech) to generate *GRB14*/pAS2.1 containing an in-frame fusion of full length Grb14 with the GAL4 DNA-BD. This construct was introduced by electroporation into the yeast strain CG1945 (*MATa, ura3-52, his3-200, ade2-101, lys2-801,trp1-901, leu2-3, 112, gal4-542, gal80-538, cyh^{r}2, LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, URA3::GAL4 17mers(x3)-CYC1$_{TATA}$-lacZ*) selecting for tryptophan prototrophy. The expression of the fusion protein was verified by Western blot analysis with antibodies directed against the Flag epitope and the Gal4 DNA-BD. The recipient strain was then grown to mid-log phase and a human liver cDNA library in the vector pACT2 (Clontech) introduced using the LiAc procedure (Schiestl and Gietz, Curr. Genet. 16, 339-346, 1989). Transformants were then selected for tryptophan, leucine and histidine prototrophy in the presence of 5mM 3-aminotriazole.

From a screen of $1x10^6$ clones, 39 colonies were initially selected on synthetic complete (SC)-leu-his-trp +3AT medium and were then tested for 9-galactosidase activity. 12 clones scored positive in the latter assay and were subjected to cycloheximide (CHX) curing to remove the bait plasmid by streaking out on SC-leu media containing 10ug/ml CHX (pAS2-1 contains the *CYH2* gene which restores CHX sensitivity to CG1945 cells). This enabled confirmation of the bait dependency of *LacZ* activation and subsequent isolation of the pACT2 plasmids encoding interacting proteins by standard methodology (Philippsen *et al*, Methods in Enzymology 194, 170-177). Back transformations were then performed in which these pACT2 plasmids were introduced into CG1945 strains containing the bait plasmid (*GRB14*/pAS2-1) or constructs encoding non-related Gal4 DNA-BD fusions in order to confirm the specificity of the interactions.

The DNA sequences of the cDNA inserts were then obtained by cycle sequencing (f-mol kit, Promega) using pACT2-specific and/or clone-specific primers. Based on their nucleotide sequences the 12 interacting clones were classified into 6 independent groups (see Table I).

TABLE I: Characterization of cDNA clones isolated by the yeast two hybrid screen.

The 12 clones exhibiting activation of both the *HIS3* and *lacZ* reporter genes were divided into 6 groups by sequence analysis of their cDNA inserts. Results of β-galactosidase activity assays performed using two methodologies are shown. The liquid culture-derived method (Galacto-Light, TROPIX) is more quantitative; results are given in mean relative light units (RLU) and are normalized for the protein content of the samples. Blue/white screening of the cDNA clones was also performed using a colony lift filter assay (Clontech). The intensity of blue colour development over approximately 2h is scored from +/- (very weak) to ++++ (strong).

Class	No. of clones	Identity	Mean RLU (Liquid assay)	Colour intensity (Filter assay)
1	6	Nedd4	$2.86x10^6$	++++
2	2	Htk	$1.86x10^5$	++
3	1	2.2412	$5.18x10^6$	++++
4	1	Proteosome	$3.88x10^2$	+/-
5	1	Somatostatin receptor	$1.45x10^3$	+/-
6	1	L-arginine:glycine amidinotransferase	$8.61x10^2$	+/-

Characterization of positive clones

Six clones were partial cDNAs corresponding to Nedd4, a multidomain protein containing a calcium-dependent phospholipid binding (CaLB) domain, four WW domains and a C-terminal region homologous to the E6-AP carboxyl-terminus (Kumar *et al*, Biochem. Biophys. Res. Commun. 185, 1155-1161, 1992; Sudol *et al* J. Biol. Chem. 270, 14733-14741, 1995; Huibregtse *et al* Proc. Natl. Acad. Sci. USA 92, 2563-2567, 1995). The latter is likely to confer E3 ubiquitin-protein ligase activity on Nedd4. The pACT2 clones isolated encoded the CaLB domain together with the first 22 amino acids of the first WW domain.

Two clones encoded the intracellular region and part of the extracellular domain of Htk, which is a RTK of the Eph family (Bennett *et al* J. Biol. Chem. 269, 14211-14218, 1994). The recruitment of Grb14 by Htk is

of interest for two reasons. First, the expression profile of both Htk and the murine homologue myk-1 are indicative of a potential role in mammary gland development and neoplasia (Andres *et al* Oncogene 9, 1461-1467, 1994; Berclaz *et al* Biochem. Biophys. Res. Comm. 226, 869-875, 1996). Second, Eph family members may be involved in the regulation of cell migration (Tessier-Lavigne, Cell 82, 345-348, 1995), which is intriguing given the homology of the Grb7 family to the *C. elegans* protein mig10 (Stein *et al*, 1994).

A novel cDNA of 1971 bp, designated 2.2412, was also isolated. This clone encoded a polypeptide of 657 amino acids in frame with the Gal4 DNA-BD (Figs. 1 and 2). The cDNA does not contain a stop codon and this, together with the Northern analysis described below, suggests that it is incomplete.

The remaining 3 clones exhibited only very weak activation of the *LacZ* reporter and were not subjected to further analysis beyond an initial characterization by DNA sequencing (Table I).

Further characterization of 2.2412

Database searches using the 2.2412 cDNA sequence revealed significant homology with a large number of proteins containing ankyrin-like repeats. These sequences were first identified as homologous regions between certain cell cycle regulatory proteins and the Drosophila protein Notch (Breeden and Nasmyth, Nature 329, 651-654, 1987) but subsequently they have been identified in a wide variety of other proteins where they are thought to function in protein-protein interactions (Bork, Proteins 17, 363-374, 1993). Subsequent analysis of the protein sequence identified 12 consecutive ankyrin repeats and an additional repetitive element (Fig. 2). The ankyrin repeat region is followed by a stretch of approximately 40 amino acids rich in serine residues. The remaining C-terminal region has a relatively high content of charged amino acids.

Northern analysis of 2.2412 mRNA expression

Northern blot analysis of multiple tissue northerns (Clontech) was performed using the 2.2412 cDNA as a probe. This resulted in the detection of a single mRNA transcript of approximately 7 kb in all tissues examined with the exception of the kidney. Expression was particularly high in

skeletal muscle and placenta. The size of this transcript compared to that of the 2.2412 clone indicates that the latter represents only a partial cDNA.

Genomic localization of the 2.2412 gene

Fluorescence *in situ* hybridization of the 2.2412 cDNA to normal metaphases (Baker *et al*, 1996) and reference to the FRA10A fragile site at 10q23.32 localized the gene to between chromosome 10q23.2 and proximal 10q23.32. Interestingly, deletions in the 10q22-25 region of chromosome 10 have been detected in a variety of human cancers including breast, prostate, renal, small cell lung and endometrial carcinomas, glioblastoma multiforme, melanoma and meningiomas, suggesting the presence of one or more tumour suppressive loci in this region (Li *et al*, Science 275, 1943-1947, 1997; Steck *et al*, Nature Genetics 15, 356-362, 1997, and references therein). Two candidate tumour suppressor genes have been identified in this region (MMAC1/PTEN and MXI1, Li *et al* 1997; Steck *et al* 1997; Albarosa *et al*, Hum. Genet. 95, 709-711, 1995).

Analysis of the interaction between 2.2412 and Grb7 family members

cDNAs encoding the full length and N- and C-terminal regions (amino acids 1-307 and 308-657, respectively) of 2.2412 were cloned into the vector pGEX4T2 (Pharmacia). The full length construct was generated by subcloning from the pACT2 clone as a NdeI fragment, whereas the shorter constructs were synthesized by directional cloning of PCR products. The corresponding GST-fusion proteins were purified from IPTG-induced bacterial cultures using glutathione-agarose beads (Smith and Johnson, Gene 67, 31-40, 1988). These immobilized fusion proteins were then incubated with lysates from cells expressing Flag epitope-tagged Grb14 (Daly *et al*, 1996) or human breast cancer cells expressing high levels of Grb7 (SK-BR-3; Stein *et al*, 1994) as described previously (Daly *et al*, 1996). Following washing, bound proteins were detected by Western blot analysis. The results indicated that 2.2412 bound specifically to both Grb14 and Grb7 *in vitro*, and that the N-terminal fusion protein bound more strongly than that derived from the C-terminus. These data, obtained using a different methodology for detecting protein-protein interactions to the yeast two hybrid system, confirm that 2.2412 interacts with Grb14. Furthermore, 2.2412 also binds

Grb7. Consequently 2.2412 appears to represent a general effector for the Grb7 family.

Grb7 family proteins exhibit differential expression in certain human cancers (particularly breast and prostate cancer) and may therefore be involved in tumour progression. Detection of the protein encoded by the cDNA 2.2412 in a sample (e.g. a homogenised tissue sample) may further provide a useful tumour marker and/or prognostic indicator for these cancers. Furthermore, the interaction of Grb7 family members with 2.2412 may provide a novel target for therapeutic intervention.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this 23rd day of September 1997.

GARVAN INSTITUTE OF MEDICAL RESEARCH
Patent Attorneys for the Applicant:

F.B. RICE & CO.

```
GAAAGATTAGCATATGAATTTAAAGGCCACTCGTTGCTGCAAGCTGCACGAGAAGCTGATGTTACTCGAA   70
 E  R  L  A  Y  E  F  K  G  H  S  L  L  Q  A  A  R  E  A  D  V  T  R
TCAAAAAACATCTCTCTCTGGAAATGGTGAATTTCAAGCATCCTCAAACACATGAAACAGCATTGCATTG  140
I  K  K  H  L  S  L  E  M  V  N  F  K  H  P  Q  T  H  E  T  A  L  H  C
TGCTGCTGCATCTCCATATCCCAAAAGAAAGCAAATATGTGAACTGTTGCTAAGAAAAGGAGCAAACATC  210
  A  A  A  S  P  Y  P  K  R  K  Q  I  C  E  L  L  L  R  K  G  A  N  I
AATGAAAAGACTAAAGAATTCTTGACTCCTCTGCACGTGGCATCTGAGAAAGCTCATAATGATGTTGTTG  280
 N  E  K  T  K  E  F  L  T  P  L  H  V  A  S  E  K  A  H  N  D  V  V
AAGTAGTGGTGAAACATGAAGCAAAGGTTAATGCTCTGGATAATCTTGGTCAGACTTCTCTACACAGAGC  350
E  V  V  V  K  H  E  A  K  V  N  A  L  D  N  L  G  Q  T  S  L  H  R  A
TGCATATTGTGGTCATCTACAAACCTGCCGCCTACTCCTGAGCTATGGGTGTGATCCTAACATTATATCC  420
  A  Y  C  G  H  L  Q  T  C  R  L  L  L  S  Y  G  C  D  P  N  I  I  S
CTTCAGGGCTTTACTGCTTTACAGATGGGAAATGAAAATGTACAGCAACTCCTCCAAGAGGGTATCTCAT  490
 L  Q  G  F  T  A  L  Q  M  G  N  E  N  V  Q  Q  L  L  Q  E  G  I  S
TAGGTAATTCAGAGGCAGACAGACAATTGCTGGAAGCTGCAAAGGCTGGAGATGTCGAAACTGTAAAAAA  560
L  G  N  S  E  A  D  R  Q  L  L  E  A  A  K  A  G  D  V  E  T  V  K  K
ACTGTGTACTGTTCAGAGTGTCAACTGCAGAGACATTGAAGGGCGTCAGTCTACACCACTTCATTTTGCA  630
  L  C  T  V  Q  S  V  N  C  R  D  I  E  G  R  Q  S  T  P  L  H  F  A
GCTGGGTATAACAGAGTGTCCGTGGTGGAATATCTGCTACAGCATGGAGCTGATGTGCATGCTAAAGATA  700
 A  G  Y  N  R  V  S  V  V  E  Y  L  L  Q  H  G  A  D  V  H  A  K  D
AAGGAGGCCTTGTACCTTTGCACAATGCATGTTCTTACGGACATTATGAAGTTGCAGAACTTCTTGTTAA  770
K  G  G  L  V  P  L  H  N  A  C  S  Y  G  H  Y  E  V  A  E  L  L  V  K
ACATGGAGCAGTAGTTAATGTAGCTGATTTATGGAAATTTACACCTTTACATGAAGCAGCAGCAAAAGGA  840
  H  G  A  V  V  N  V  A  D  L  W  K  F  T  P  L  H  E  A  A  A  K  G
AAATATGAAATTTGCAAACTTCTGCTCCAGCATGGTGCAGACCCTACAAAAAAAAACAGGGATGGAAATA  910
 K  Y  E  I  C  K  L  L  L  Q  H  G  A  D  P  T  K  K  N  R  D  G  N
CTCCTTTGGATCTTGTTAAAGATGGAGATACAGATATTCAAGATCTGCTTAGGGGAGATGCAGCTTTGCT  980
T  P  L  D  L  V  K  D  G  D  T  D  I  Q  D  L  L  R  G  D  A  A  L  L
AGATGCTGCCAAGAAGGGTTGTTTAGCCAGAGTGAAGAAGTTGTCTTCTCCTGATAATGTAAATTGCCGC 1050
  D  A  A  K  K  G  C  L  A  R  V  K  K  L  S  S  P  D  N  V  N  C  R
GATACCCAAGGCAGACATTCAACACCTTTACATTTAGCAGCTGGTTATAATAATTTAGAAGTTGCAGAGT 1120
 D  T  Q  G  R  H  S  T  P  L  H  L  A  A  G  Y  N  N  L  E  V  A  E
ATTTGTTACAACACGGAGCTGATGTGAATGCCCAAGACAAAGGAGGACTTATTCCTTTACATAATGCAGC 1190
Y  L  L  Q  H  G  A  D  V  N  A  Q  D  K  G  G  L  I  P  L  H  N  A  A
ATCTTACGGGCATGTAGATGTAGCAGCTCTACTAATAAAGTATAATGCATCTCTCAATGCCACGGACAAA 1260
  S  Y  G  H  V  D  V  A  A  L  L  I  K  Y  N  A  S  L  N  A  T  D  K
TGGGCTTTCACACCTTTGCACGAAGCAGCCCAAAAGGGACGAACACAGCTTTGTGCTTTGTTGCTAGCCC 1330
 W  A  F  T  P  L  H  E  A  A  Q  K  G  R  T  Q  L  C  A  L  L  L  A
ATGGAGCTGACCCGACTCTTAAAAATCAGGAAGGACAAACACCTTTAGATTTAGTTTCAGCAGATGATGT 1400
H  G  A  D  P  T  L  K  N  Q  E  G  Q  T  P  L  D  L  V  S  A  D  D  V
CAGCGCTCTTCTGACAGCAGCCATGCCCCCATCTGCTCTGCCCTCTTGTTACAAGCCTCAAGTGCTCAAT 1470
  S  A  L  L  T  A  A  M  P  P  S  A  L  P  S  C  Y  K  P  Q  V  L  N
GGTGTGAGAAGCCCAGGAGCCACTGCAGATGCTCTCTCTTCAGGTCCATCTAGCCCATCAAGCCTTTCTG 1540
 G  V  R  S  P  G  A  T  A  D  A  L  S  S  G  P  S  S  P  S  S  L  S
CAGCCAGCAGTCTTGACAACTTATCTGGGAGTTTTTCAGAACTGTCTTCAGTAGTTAGTTCAAGTGGAAC 1610
A  A  S  S  L  D  N  L  S  G  S  F  S  E  L  S  S  V  V  S  S  S  G  T
AGAGGGTGCTTCCAGTTTGGAGAAAAAGGAGGTTCCAGGAGTAGATTTTAGCATAACTCAATTCGTAAGG 1680
  E  G  A  S  S  L  E  K  K  E  V  P  G  V  D  F  S  I  T  Q  F  V  R
AATCTTGGACTTGAGCACCTAATGGATATATTTGAGAGAGAACAGATCACTTTGGATGTATTAGTTGAGA 1750
 N  L  G  L  E  H  L  M  D  I  F  E  R  E  Q  I  T  L  D  V  L  V  E
TGGGGCACAAGGAGCTGAAGGAGATTGGAATCAATGCTTATGGACATAGGCACAAACTAATTAAAGGAGT 1820
M  G  H  K  E  L  K  E  I  G  I  N  A  Y  G  H  R  H  K  L  I  K  G  V
CGAGAGACTTATCTCCGGACAACAAGGTCTTAACCCATATTTAACTTTGAACACCTCTGGTAGTGGAACA 1890
  E  R  L  I  S  G  Q  Q  G  L  N  P  Y  L  T  L  N  T  S  G  S  G  T
ATTCTTATAGATCTGTCTCCTGATGATAAAGAGTTTCAGTCTGTGGAGGAAGAGATGCAAAGTACAGTTC 1960
 I  L  I  D  L  S  P  D  D  K  E  F  Q  S  V  E  E  E  M  Q  S  T  V
GAGAGCACAGA                                                              1971
R  E  H  R
```

Fig. 1

1 ERLAYEFKGHSLLQAAREADVTRIKKHLSLEMVNFKHPQ**THETALHCAAASPYPKRKQICELLL**

65 **RKGANINE**KTK**EFLTPLHVASEKAHNDVVEVVVKHEAKVNA**LDN**LGQTSLHRAAYCGHLQTCRL**

129 **LLSYGCDPNIISLQGFTALQMGNENVQQLLQEGISLGNSEADRQLL**EAAKAGDVETVKKLCTVQ

193 SVNCRDIEG**RQSTPLHFAAGYNRVSVVEYLLQHGADVHAKDKGGLVPLHNACSYGHYEVAELLV**

257 **KHGAVVNVADLWKFTPLHEAAAKGKYEICKLLLQHGADPTKKNRDGNTPLDLVKDGDTDIQDLL**

321 **RGDAALLDAAKKG**CLARVKKLSSPDNVNCRDTQG**RHSTPLHLAAGYNNLEVAEYLLQHGADVNA**

385 **QDKGGLIPLHNAASYGHVDVAALLIKYNASLNATDKWAFTPLHEAAQKGRTQLCALLLAHGADP**

449 **TLKNQEGQTPLDLVSADDVSALLTAAMPPSALPSCYKP**QVLNGVRSPGATADALSSGPSSPSSL

513 SAASSLDNLSGSFSELSSVVSSSGTEGASSLEKKEVPGVDFSITQFVRNLGLEHLMDIFEREQI

577 TLDVLVEMGHKELKEIGINAYGHRHKLIKGVERLISGQQGLNPYLTLNTSGSGTILIDLSPDDK

641 EFQSVEEEMQSTVREHR

Fig 2.